

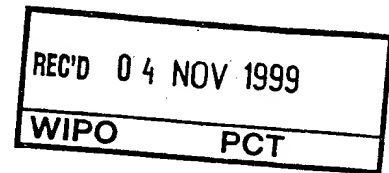
09/807638

**PRIORITY  
DOCUMENT**

SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)



**Bescheinigung**



Die VOLKSWAGEN Aktiengesellschaft in Wolfsburg/Deutschland hat eine  
Patentanmeldung unter der Bezeichnung

"Verfahren und Vorrichtung zur automatischen Steuerung  
mindestens eines Gerätes per Sprachdialog"

am 25. Februar 1999 beim Deutschen Patent- und Markenamt eingereicht  
und erklärt, daß sie dafür die Innere Priorität der Anmeldung in der Bundesrepublik  
Deutschland vom 16. Oktober 1998, Aktenzeichen 198 47 879.8, in Anspruch  
nimmt.

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprüngli-  
chen Unterlagen dieser Patentanmeldung.

Die Anmeldung hat im Deutschen Patent- und Markenamt vorläufig das Symbol  
G 10 L 15/00 der Internationalen Patentklassifikation erhalten.

München, den 6. Oktober 1999

**Deutsches Patent- und Markenamt**

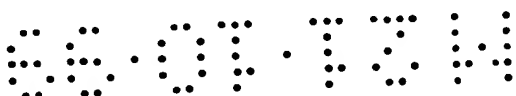
**Der Präsident**

Im Auftrag

Joost

Aktenzeichen: 199 08 137.9

VOLKSWAGEN



K 7341IP/1770-hk-sk

## Verfahren und Vorrichtung zur automatischen Steuerung mindestens eines Gerätes per Sprachdialog

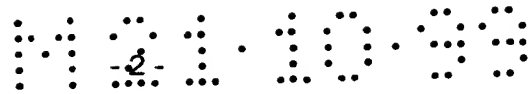
Die Erfindung betrifft ein Verfahren und eine Vorrichtung zur automatischen Steuerung mindestens eines Gerätes per Sprachdialog.

Verfahren bzw. Vorrichtungen dieser Art werden allgemein in sogenannten Sprachdialog- bzw. Sprachbediensystemen z.B. für Fahrzeuge, Computer, Roboter, Maschinen, Anlagen usw. eingesetzt.

Ein Sprachdialogsystem (SDS) läßt sich i.a. im wesentlichen auf folgende Komponenten reduzieren:

- Spracherkennungssystem, welches ein eingesprochenes Kommando ("Sprachkommando") mit anderen erlaubten Sprachkommandos vergleicht und eine Entscheidung trifft, welches Kommando aller Wahrscheinlichkeit nach ausgesprochen wurde,
- Sprachausgabe, welche die zur Benutzerführung erforderlichen Sprachbefehle und Signalisierungstöne ausgibt und ggf. das Erkennungsergebnis rückmeldet,
- Dialog- und Ablaufsteuerung, um dem Benutzer zu verdeutlichen, welche Art von Eingabe erwartet wird, bzw. um zu prüfen, ob die erfolgte Eingabe konsistent ist mit der Anforderung und mit dem momentanen Status der Applikation, und um die resultierende Aktion bei der Applikation (z.B. dem zu steuernden Gerät) anzustoßen, Kontrollinterface als Schnittstelle zur Applikation: Dahinter verbergen sich Hard- und Softwaremodule, um verschiedene Aktuatoren bzw. Rechner anzusteuern, die die Applikation beinhalten,
- Applikation, die per Sprache angesteuert wird: Das kann z.B. ein Bestell- oder Auskunftssystem, ein CAE-Arbeitsplatz oder ein behindertengerechter Rollstuhl sein.

Ein solches Sprach-Dialog-System ist beispielsweise aus der DE 195 33 541 C1 bekannt. Zur Verbesserung der Akzeptanz eines solchen Mensch-Maschinen-Dialogs wird dort beispielsweise vorgeschlagen, für die Kommandos synonyme Worte bzw. unterschiedliche Aussprachevarianten zu verwenden oder aber die Wörter in den Kommandos umzustellen, beispielsweise "Größerer Radius bei linkem Kreis" oder alternativ hierzu "Bei linkem Kreis



größerer Radius“. Des weiteren wird vorgeschlagen, durch einen erweiterten Speicher ein multilinguales sprecherunabhängiges Dialogsystem aufzubauen, wobei wahlweise zwischen den Dialogsystemen verschiedener Sprachen umgeschaltet werden kann. Weiter wird vorgeschlagen, sogenannte Ellipsen einzubinden, d.h. Verzicht auf Wiederholung ganzer Kommandosätze und statt dessen Kommandos wie „höher“, „stärker“ oder „weiter“, wobei dann das Sprachdialog-System diese den vorangegangenen Kommandos zuordnet. Des weiteren kann das Sprach-Dialog-System bei unsicherem Erkennen Rückfragen wie beispielsweise „Wie bitte?“, „Bitte wiederholen“ oder „Und weiter?“ stellen oder bestimmte Anregungen wie „Bitte lauter“ ausgeben. All diese Maßnahmen dienen dazu, eine monotone Kommunikation zu vermeiden und den Dialog stärker einer Mensch-Mensch-Kommunikation anzunähern. Zur Verbesserung der Kommunikation wird das Sprachsystem mit einem optischen Anzeigemedium gekoppelt, auf dem erkannte Befehle zu Kontrollzwecken angegeben werden. Weiter ermöglicht das optische Anzeigemedium die Darstellung der vom Zielgerät als Reaktion auf das Sprachkommando eingestellten Funktionen und/oder verschiedener Funktionen/Alternativen, die per Sprachkommando anschließend eingestellt bzw. ausgewählt werden können. Nachteilig an der bekannten Vorrichtung und des damit durchgeführten Verfahrens ist, daß trotz der dargestellten Verbesserungen die Sprachausgabe aufgrund ihrer Monotonie den Nutzer ermüdet, so daß dessen Reaktionszeit bei handlungskritischen Aktionen zu groß ist. Ein weiteres Problem ist, daß bei Erkennungsschwierigkeiten die bekannten Sprach-Dialog-Systeme eine Endlosschleife durchlaufen und dem Nutzer immer wieder die gleiche Aufforderung ausgeben, so daß der work-flow unterbrochen ist.

Der Erfindung liegt daher das technische Problem zugrunde, ein Verfahren und eine Vorrichtung zur automatischen Steuerung eines Gerätes per Sprachdialog zu schaffen, bei denen bei handlungskritischen Anweisungen eine verbesserte Aufmerksamkeit des Nutzers erreicht wird sowie der work-flow sichergestellt ist.

Die Lösung des technischen Problems ergibt sich durch die Merkmale der Patentansprüche 1, 2 und 8. Weitere vorteilhafte Ausgestaltungen der Erfindung ergeben sich aus den Unteransprüchen.

Durch die Verwendung unterschiedlicher Intonationen, wobei bei handlungskritischen Anweisungen die Kommandos eine Befehlsintonation aufweisen, wird bereits unmittelbar bei der Sprachausgabe die Aufmerksamkeit des Nutzers geweckt, so daß die Reaktionszeit zur Durchführung der geforderten Anweisung erheblich verkürzt wird.

Zur weiteren Erhöhung der Aufmerksamkeit und Unterscheidung von handlungskritischen Anweisungen kann die Lautstärke der Sprachausgabe bei handlungskritischen Anweisungen erhöht werden und/oder diese besonders hart bzw. abrupt eingeblendet werden.

Weiter kann das Sprach-Dialog-System multilingual ausgebildet sein, so daß beispielsweise zwischen einer Männer- und einer Frauenstimme gewählt werden kann, wobei dann bei handlungskritischen Anweisungen die eine und bei handlungsunkritischen Anweisungen die andere Stimme vom System ausgewählt wird.

Zur Sicherstellung des work-flows wird das Sprach-Dialog-System nur durch Betätigung eines „Push to talk“ PTT-Schalters aktiviert, wobei bei Ausbleiben einer gültigen Interaktion die Dialog-Kommunikations-Ebene gewechselt wird. Zur Erhöhung der Erkennungssicherheit bzw. zur Verbesserung der Nutzerführung können einzelne Kommandos in verschiedenen alternativen Ausgabeformen abgespeichert sein, die dann bei einer ungültigen Interaktion sukzessive ausgegeben werden, wobei die Dialog-Kommunikations-Ebene erst gewechselt wird, wenn auf alle Kommandoformen keine gültige Interaktion erfolgt ist. Zur Vermeidung von Monotonie kann die Reihenfolge der Ausgabe durch einen Zufallsgenerator permutiert werden.

Der Grundgedanke der Erfindung besteht darin, durch die Art der Sprachausgabe beim Kraftfahrzeugführer eine Emotion zu wecken, die ein situationsgerechtes Handeln provoziert.

Die Erfindung wird nachfolgend anhand eines bevorzugten Ausführungsbeispiels näher erläutert. Die Fig. zeigen:

Fig. 1 ein schematisches Flußdiagramm des Verfahrens zur automatischen Steuerung mindestens eines Gerätes per Sprachdialog,

Fig. 2a einen schematischen Verlauf des Gefahrenpotentials bei handlungsunkritischen Interaktionen,

Fig. 2b einen zugehörigen Denotationsverlauf,

Fig. 2c einen zugehörigen Intonationsverlauf,

Fig. 2d einen zugehörigen Konnotationsverlauf,

Fig. 3a einen schematischen Verlauf des Gefahrenpotentials bei handlungskritischen Interaktionen,

Fig. 3b einen zugehörigen Denotationsverlauf,

Fig. 3c einen zugehörigen Intonationsverlauf und

Fig. 3d einen zugehörigen Konnotationsverlauf.

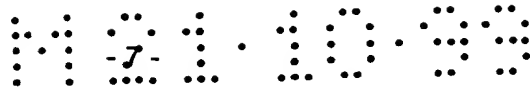
In den Fig. 2a-d sind schematisch die Verhältnisse für eine handlungsunkritische Anweisung, wie beispielsweise eine Angabeanforderung für ein Navigationssystem, dargestellt. Dabei ist in Fig. 2a die Wichtigkeit der Interaktion über die Zeit  $t$  aufgetragen, wobei zu den Zeitpunkten  $t_0 - t_2$  Handlungsanweisungen ausgegeben werden, wobei angenommen wird, daß auf jede vorangegangenen Aufforderung nicht reagiert wurde. Da eine fehlende Eingabe in das Navigationssystem nur dazu führt, daß diese Komfortkomponente nicht arbeiten kann, was gegebenenfalls auch nicht vom Kraftfahrzeugführer erwünscht ist, ändert sich die Wichtigkeit über die Zeit nicht. Ebenso bleibt der die Inhaltsangabe des Befehls, die sogenannte Denotation, nämlich die Eingabeanforderung über die Zeit konstant, was in Fig. 2b dargestellt ist. Zum Zeitpunkt  $t_0$  wird nun beispielsweise der Kraftfahrzeugführer aufgefordert, „Bitte geben Sie nun Ihr gewünschtes Fahrziel ein“. Diese Aufforderung erfolgt mit einer bestimmten Intonation  $I_1$  und einem bestimmten Begriffsinhalt (Konnotat)  $K1$ , was in Fig. 2c-d dargestellt ist. Erfolgt nun keine Eingabe, so weiß das System nicht, warum dies unterblieben ist, beispielsweise ob der Kraftfahrzeugführer die Aufforderung überhört hat oder willentlich nicht vorzunehmen gedenkt. Daher erfolgt zum Zeitpunkt  $t_1$  eine erneute Aufforderung „Wollen Sie ein Fahrziel eingeben“ mit einer verstärkten Intonation  $I_2$ , um eine verbesserte Wahrnehmbarkeit zu erreichen, wobei jedoch die Konnotationsstufe  $K2$  abnimmt. Erfolgt hierauf wieder keine Eingabe, so kann das System davon ausgehen, daß dies der Kraftfahrzeugführer wohl nicht wünscht. Um die Belästigung des Kraftfahrzeugführeres durch permanente Wiederholungen gering zu halten, erfolgt dann zum Zeitpunkt  $t_2$  eine letztmalige Aufforderung, wie beispielsweise „Falls Sie keinen Zielort wünschen, schalte ich mich jetzt ab“. Diese letzte Aufforderung wird mit einer sehr geringen Intonation  $I_3$  ausgegeben und weist nur noch eine geringe Konnotation auf. Wie aus dem Verlauf in Fig. 2d zu entnehmen ist, bildet die Konnotation einen Antiklimax, also einen Übergang vom stärkeren zum schwächeren Ausdruck, wohingegen bei der Intonation eine gewisse Variation auftritt, um Monotonie entgegenzuwirken.

Im Gegensatz hierzu ist in Fig. 3a-d eine Situation dargestellt, bei der Wichtigkeit der Interaktion mit der Zeit zunimmt und schließlich handlungskritisch wird. Beispielsweise fährt das Kraftfahrzeug mit einer Geschwindigkeit oberhalb einer zugelassenen Geschwindigkeit auf der Autobahn unter Einhaltung des Sicherheitsabstandes hinter einem Kraftfahrzeug hinterher. Zum Zeitpunkt  $t_0$  gibt das System eine Handlungsanweisung, beispielsweise der Form „Bitte passen Sie Ihre Geschwindigkeit an“, an den Kraftfahrzeugführer. Die Handlungsanweisung weist einen geringen Intonationsgrad  $I_1$  und eine entsprechend geringe Konnotationsstufe  $K1$  auf, da der Kraftfahrzeugführer zwar eine Ordnungswidrigkeit begeht, jedoch keine unmittelbare Gefahr besteht. Weiter sei angenommen, daß nun der Kraftfahrzeugführer seine Geschwindigkeit nicht anpaßt und zusätzlich zum Zeitpunkt  $t_1$  der

Durch die Betätigung des PTT-Schalters wird das Sprach-Dialog-System aktiviert. Zum besseren Verständnis wird die Sprachausgabe des Sprachdialog-Systems in Kommandos KOM und Aufforderungen Auff unterteilt, die in Realität identisch sein können. Nachfolgend wird unter Kommandos KOM eine direkte Anweisung zum Handeln verstanden, wie beispielsweise „BREMSEN“, „LICHTEINSCHALTEN“, wohingegen Aufforderungen eine Interaktion in Form einer Eingabe anfordern, wie beispielsweise „Bitte gewünschte Temperatur in Grad Celsius angeben“.

Erzeugt nun das Sprach-Dialog-System ein Kommando KOM, so wird dieses Kommando KOM danach unterteilt, ob es sich um eine handlungskritische Anweisung oder eine handlungsunkritische Anweisung handelt. Vereinfacht sind handlungskritische Anweisungen, Kommandos KOM, die ein schnelles Durchführen der Handlung erfordern. Ein Beispiel hierfür ist das Kommando KOM „Bremsen“, wenn ein ADR-System oder eine Prescrash-Sensorik ein Kollisionsobjekt erfaßt haben. Handlungsunkritische Anweisungen sind beispielsweise Kommandos KOM eines Navigationssystems. Dabei werden handlungskritische Anweisungen mit einer Befehlsintonations-Sprache S1, mit großer Lautstärke L1 hart und abrupt in einer Zeit t1 eingeblendet, um beim Nutzer einen hohen Aufmerksamkeitsgrad zu erzeugen. Handlungsunkritische Anweisungen werden hingegen weich bei geringerer Lautstärke L2 und normaler Intonation S2 eingeblendet.

Aufforderungen Auff sind in der Regel zeitunkritisch, so daß hier ein gute Benutzerführung im Vordergrund steht. Dazu kann eine Aufforderung Auff in n verschiedenen Alternativen im Sprachspeicher abgelegt sein. Die Alternativen können beispielsweise unterschiedliche Betonungen, Aussprachen, Wortumstellungen oder synonyme Begriffe sein. Nach akustischer Ausgabe der ersten Alternative wartet das Sprach-Dialog-System eine vorgegebene Zeit auf eine Interaktion. Erfolgt innerhalb dieser Zeit keine oder eine ungültige Interaktion, so wiederholt das Sprach-Dialog-System die Aufforderung mit der zweiten Alternative, gegebenenfalls bis zur n-ten Alternative. Erfolgt eine gültige Interaktion, so wird diese ausgeführt und gegebenenfalls eine neue Aufforderung Auff ausgegeben. Erfolgt hingegen auf die n-te Alternative der Aufforderung Auff keine gültige Interaktion, so wird auf eine andere Dialog-Kommunikations-Ebene DKE gewechselt, um den work-flow sicherzustellen. Die neue Dialog-Kommunikations-Ebene DKE ist dann beispielsweise eine Auswahlliste, die auf dem Bordmonitor dargestellt wird, woraus sich der Nutzer ein entsprechendes Menü auswählen kann.



K 7341IP/1770-hk-sk

## PATENTANSPRÜCHE

1. Vorrichtung zur automatischen Steuerung mindestens eines Gerätes per Sprachdialog, umfassend einen Sprachspeicher, in dem für die Sprachanlage benötigte Kommandos abgelegt sind,  
dadurch gekennzeichnet, daß  
die Kommandos (KOM) in unterschiedlichen Intonationen abgelegt sind, wobei mindestens die Kommandos (KOM) für handlungskritische Anweisungen eine Befehlsintonation aufweisen.
2. Verfahren zur automatischen Steuerung mindestens eines Gerätes per Sprachdialog, mittels mindestens eines Sprachspeichers, in dem für die Sprachausgabe benötigte Kommandos (KOM) abgelegt sind,  
dadurch gekennzeichnet, daß  
bei handlungskritischen Anweisungen die Kommandos (KOM) mit Befehlsintonation ausgegeben werden.
3. Verfahren nach Anspruch 2,  
dadurch gekennzeichnet, daß  
die handlungskritischen Kommandos (KOM) mit einer höheren Lautstärke (L1) ausgegeben werden.
4. Verfahren nach Anspruch 2 oder 3,  
dadurch gekennzeichnet, daß  
die handlungskritischen Kommandos (KOM) hart eingeblendet werden.
5. Verfahren nach einem der Ansprüche 2 bis 4,  
dadurch gekennzeichnet, daß  
die Kommandos (KOM) multilingual im Sprachspeicher abgelegt sind und bei handlungskritischen Kommandos (KOM) die Sprache gewechselt wird.

Sicherheitsabstand bereits knapp unterschritten wird, d.h. das Gefahrenpotential der Verkehrssituation nimmt zu, was durch die ansteigende Kurve in Fig. 3a symbolisiert wird. Das System gibt daraufhin eine Handlungsanweisung, beispielsweise der Form „Sie müssen bremsen“ oder „Bitte bremsen Sie“ an den Kraftfahrzeugführer, wobei diese einen höheren Intonationsgrad  $I_2$  mit entsprechend höherer Konnotationsstufe  $K_2$  aufweist. Reagiert der Kraftfahrzeugführer auch hierauf nicht, so wird das Gefahrenpotential der Verkehrssituation noch gesteigert, was anhand des weiteren Anstiegs in Fig. 3a dargestellt ist. Dies bedeutet, daß ein weiteres Nicht-Reagieren des Kraftfahrzeugführers innerhalb kürzester Zeit zu einem Unfall führen könnte. Diese handlungskritische Anweisung wird beispielsweise in der Form „Vollbremsung“ mit einer Befehlsintonation  $I_3$  ausgegeben. Die in Fig. 3d dargestellten Konnotationsstufen stellen dabei einen Klimax dar, also die Steigerung des Ausdrucks vom weniger Wichtigen zum Wichtigeren. Ergänzend sei angemerkt, daß die in Fig. 2a-d und Fig. 3a-d dargestellten Veränderungen nicht maßstäblich sind, sondern nur als qualitative Angaben zu verstehen sind.



M21.10.99

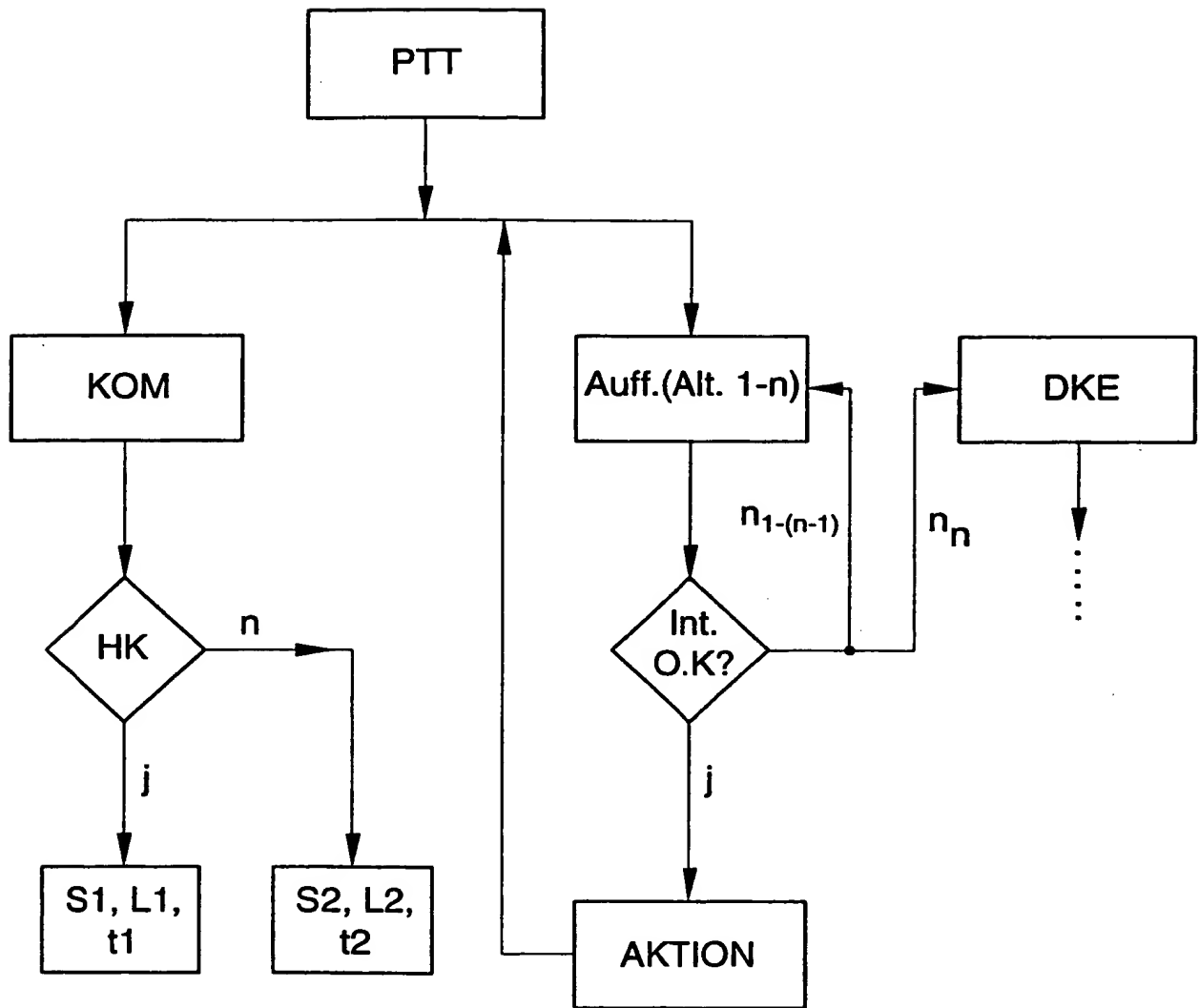


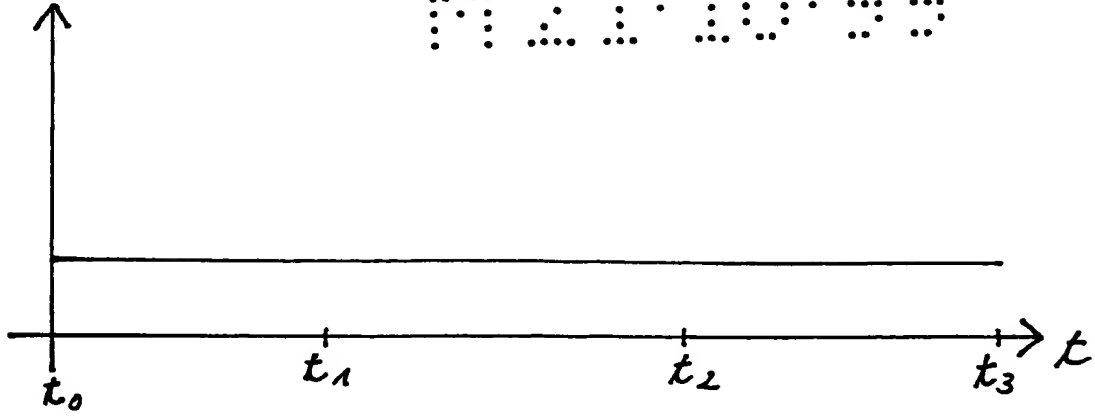
FIG. 1

6. Verfahren nach einem der Ansprüche 2 bis 5,  
dadurch gekennzeichnet, daß  
bei handlungskritischen Kommandos (KOM) die Intonation I und die Konnotation K  
entsprechend der Wichtigkeit gesteigert wird.
7. Verfahren nach einem der Ansprüche 2 bis 6,  
dadurch gekennzeichnet, daß  
bei handlungunkritischen Aufforderungen die Intonation bei abnehmender Konnotation  
K variiert wird.
8. Verfahren zur automatischen Steuerung mindestens eines Gerätes per Sprachdialog,  
mittels mindestens eines PTT-Schalters , eines Bordmonitor und eines Sprach-Dialog-  
Systems, umfassend folgende Verfahrensschritte :
  - a) Einleiten des Verfahrens durch Betätigung des PTT-Schalters,
  - b) Darstellen der möglichen Steuermöglichkeiten alphanumerisch in der  
Kommandoform und grafisches Hervorheben der sprachaktiven Funktionen und
  - c) Wechseln der Dialog-Kommunikations-Ebene, falls keine erwartete Interaktion  
erfolgt.

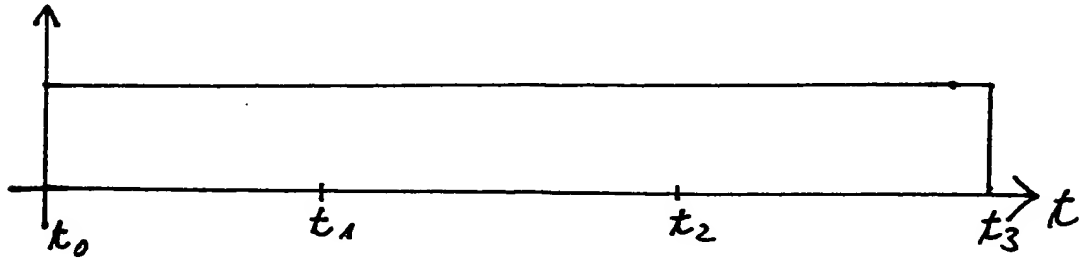
Fig 2

11.21.10.99

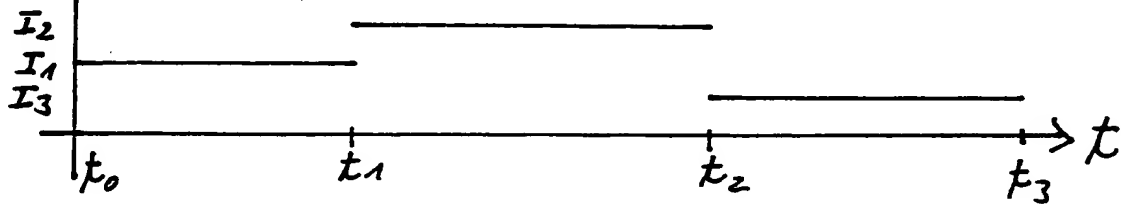
a)



b)



c)  $I \uparrow$



d)

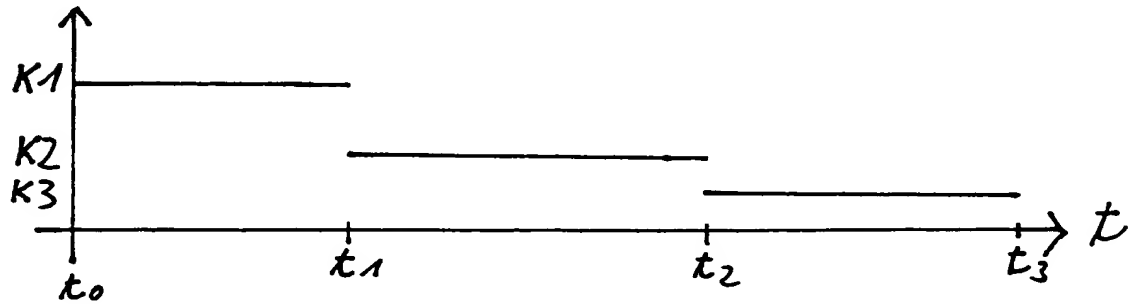
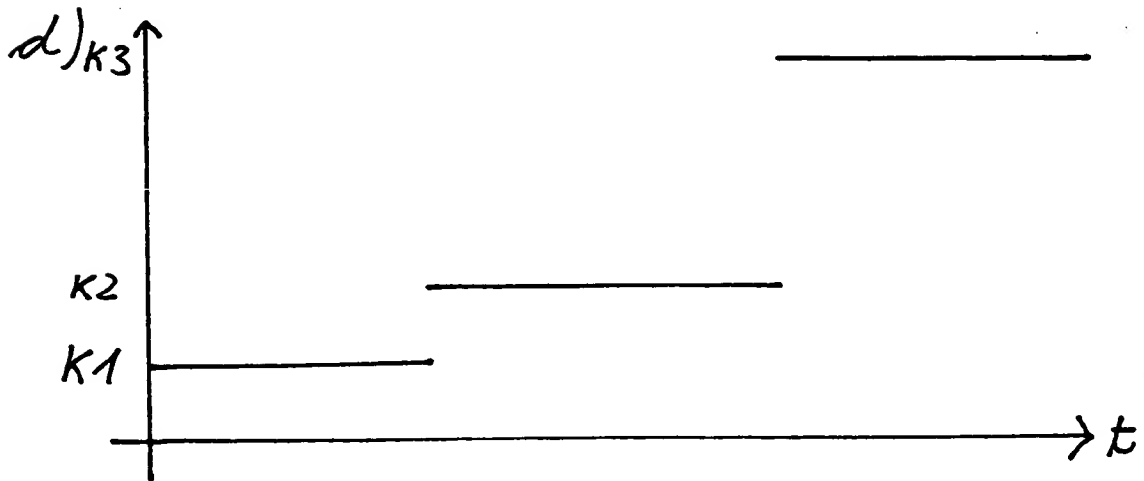
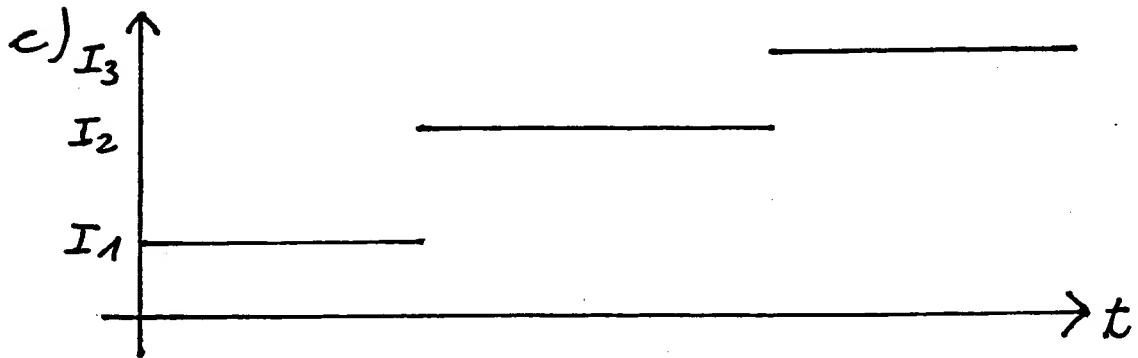
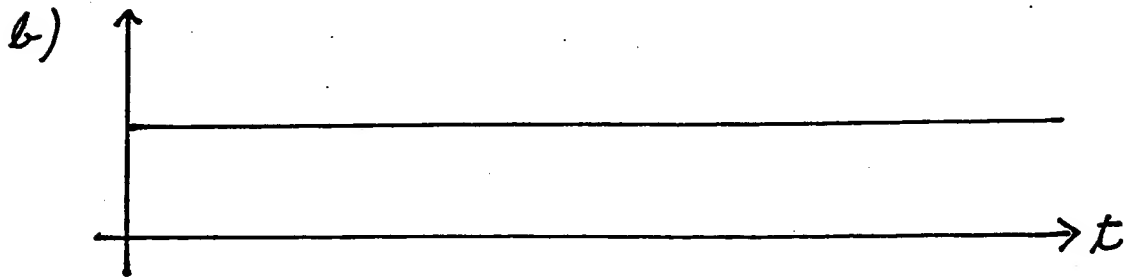
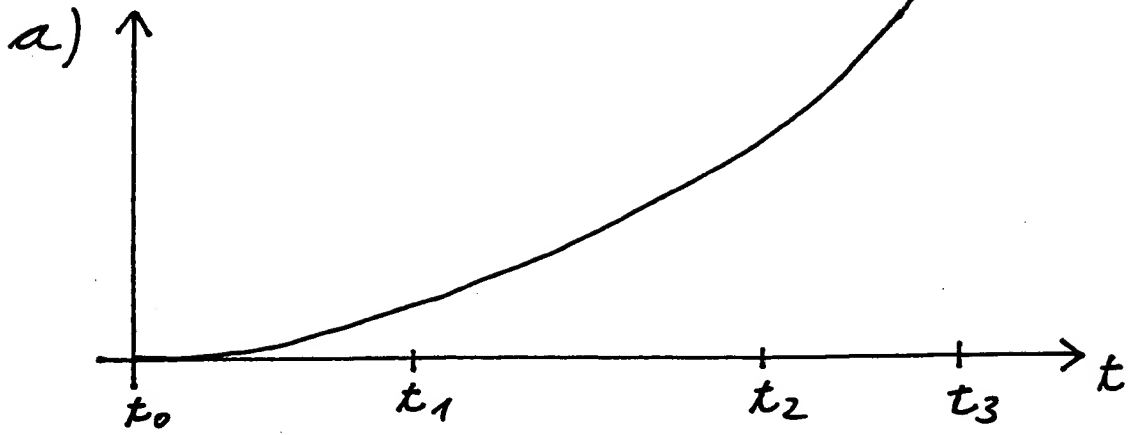


Fig. 3

11.10.99



Pat. 10.99

K 7341IP/1770-hk-sk

## **ZUSAMMENFASSUNG**

### **Verfahren und Vorrichtung zur automatischen Steuerung mindestens eines Gerätes per Sprachdialog**

Die Erfindung betrifft ein Verfahren und eine Vorrichtung zur automatischen Steuerung mindestens eines Gerätes per Sprachdialog, umfassend einen Sprachspeicher, in dem für die Sprachausgabe benötigte Kommandos (KOM) in unterschiedlichen Intonationen abgelegt sind, wobei mindestens die Kommandos (KOM) für handlungskritische Anweisungen eine Befehlsintonation aufweisen.

(Fig. 1)

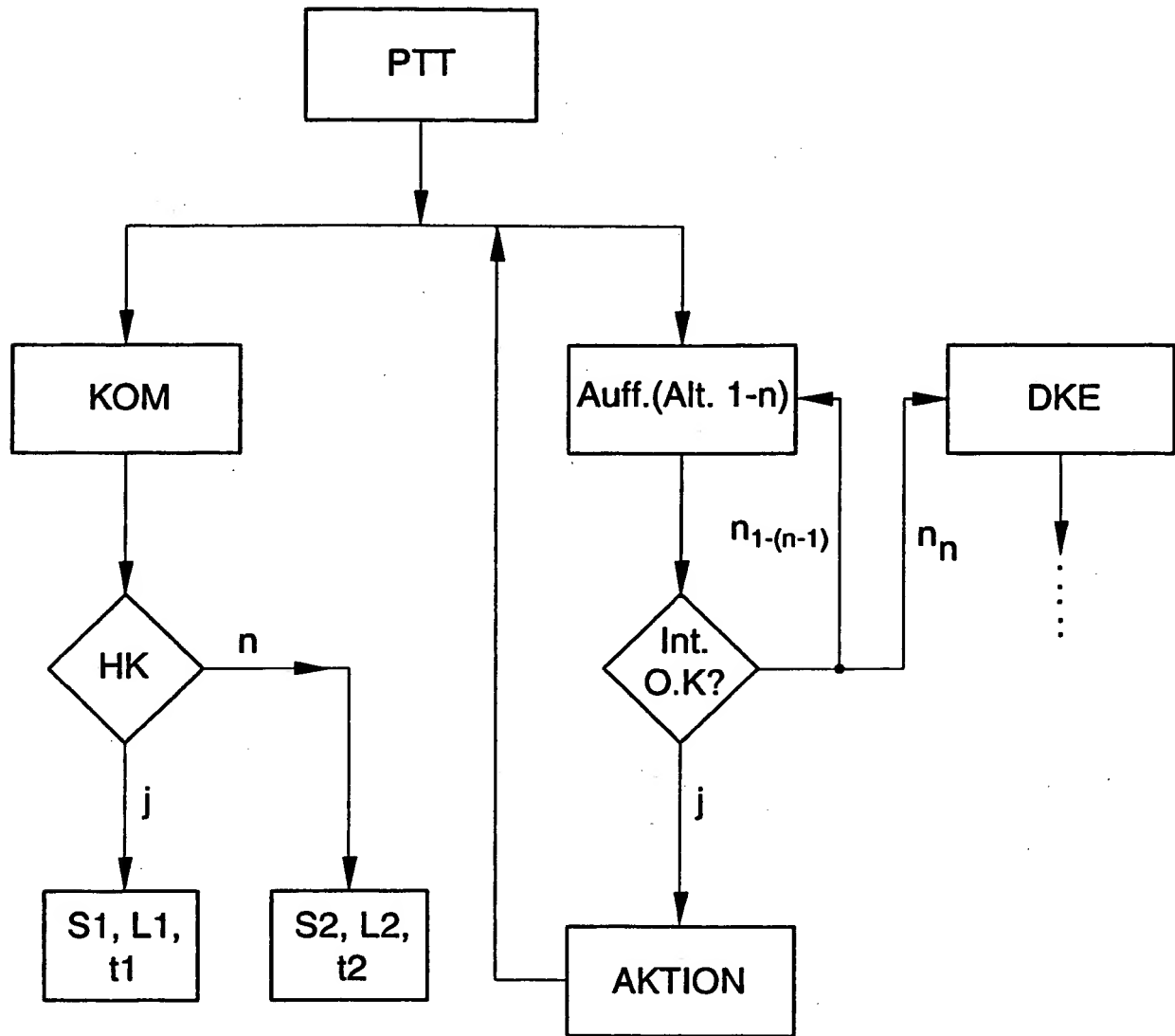


FIG. 1